

DATA PROCESSING

GOALS AND STANDARDS

Goals

- Ensure standard processing procedures are performed on data across sites.
- Process data quickly to ensure timely information is released.

Standards

- Ensure editing procedures are used to remove inconsistent records from the data collection cases.
- Ensure calculated variables are analyzed consistently across participating sites.

DATA PROCESSING

IMPORTANT CONSIDERATIONS

Additional weighting protocol When data are used without weights, each record counts the same as any other record. Implicit in such use is the assumption that each record has an equal probability of selection and that noncoverage and nonresponse are equal among all segments of the population. When the sample design results in records with different probabilities of selection and when noncoverage and nonresponse are not equal among all segments of a population, then weighting each record differently can adjust for these factors. A conceptually unrelated reason for weighting is to make the total number of cases equal to some desired population number. Poststratification can serve as a blanket adjustment for noncoverage, and nonresponse and forces the total number of cases to equal the population estimates.

DATA PROCESSING

ACTION STEPS

The factor is then multiplied by the raw weight to compute an adjusted, or final-weight, variable. Weighting of the sample adjusts not only for variation in selection and sampling probability, but also for demographic characteristics so that projections can be made from the sample to the general population. Weighting also adjusts for nonresponse and noncoverage (i.e., failure of some segments to be included in the sampling frame).

Calculate risk factor variables

Combine responses across various questions to create a set of standardized risk factors that form the basis of the surveillance system tabulations. An example of a risk factor variable is "no leisure-time physical activity," which is based on a combination of responses to questions on participation in exercise, recreation, or physical activities other than regular job (such as calisthenics, golf, gardening, or walking for exercise).

Produce frequencies of variables

After files have been edited and weighted, generate frequencies for health risk and demographic variables.

Produce edit reports

Edit reports should be generated to include the following information:

- Inconsistencies found in incomplete records, or item nonresponse
- Inconsistencies found in complete records, such as conflicting data from two or more questions, or questions for which the response is outside an acceptable range

DATA PROCESSING

ACTION STEPS

Edit the data Edits are designed to remove inconsistencies from a data set. The types of edits that can be done include:

- Checking for record errors to be sure that all IDs are valid and that there are no duplicate IDs
- Checking ranges to be certain that each field contains a valid code
- Checking one field against another within a record to be certain that the values coded in each field are consistent with one another (e.g., a respondent who indicates that he does not smoke, yet also indicates that he smokes 10 cigarettes per day)
- Checking the logical progression through a questionnaire, including use of skip patterns, based on predetermined responses

Weight the data Add weighting factors to each record to provide unbiased, representative prevalence estimates. Weighting compensates for unequal selection probabilities and nonresponse differences (i.e., their overrepresentation or underrepresentation) in the sample. Final weighting adjusts for several factors:

- Number of adults per household
- Number of interviews completed per household
- Poststratification by region (city, area) and population distribution according to age, race, and sex

Number of adults and number of interviews address the problem of unequal selection probability, which could result in a biased sample (i.e., one that does not fairly represent the population). For example, a respondent in a one-adult household has four times the chance of being selected for an interview as does a respondent in a four-adult household.

Overrepresentation or underrepresentation of any single record is addressed through poststratification. This method adjusts the distribution of the sample data so that, collectively, it reflects the total population of the sampled area. The poststratification factor is calculated by computing the ratio of the age, race, and sex distribution of the country population divided by that of the sample.

DATA PROCESSING

Before data can be analyzed, they must be processed to ensure that they are the highest quality data possible.

Data processing **Action Steps** include:

1. Edit the data
2. Weight the data
3. Calculate risk factor variables
4. Produce frequencies of variables
5. Produce edit reports

Important Consideration for data processing include:

1. Additional weighting protocol

Goals and standards include:

Goals:

1. Ensure standard processing procedures are performed on data across sites.
2. Process data quickly to ensure timely information is released.

Standards:

1. Ensure editing procedures are used to remove inconsistent records from the data collection cases.
2. Ensure calculated variables are analyzed consistently across participating sites.